UTILIZING CLASSICAL SCALING FOR FAULT IDENTIFICATION BASED ON CONTINUOUS-BASED PROCESS

MIRA SYAHIRAH BT ABD WAHIT

Thesis submitted in fulfillment of the requirements
for the degree of
Bachelor Of Chemical Engineering

Faculty of Chemical Engineering and Chemical Resources
UNIVERSITI MALAYSIA PAHANG

JANUARY 2013

ABSTRACT

This study is about to develop a new method on Fault Identification using Classical Scaling. Process Monitoring is from Statistical Process Control (SPC), the statistical tool which used of statistical methods and to control of a process, by repeated sampling measurements or to predict results. It also help to determine whether the process is working properly or not. This Statistical Process Control (SPC) will forms charts with data (control charts) to shown the result. This study will develop the normal Multivariate Statistical Process Monitoring (MSPM). The data will be run as the normal Multivariate Statistical Process Monitoring (MSPM). So from the technique that used in this study such as Principal Component Analysis, Statistical Process Control and Multivariate Statistical Process Monitoring (MSPM), this study will be develop which the most faster technique that will be detecting the fault in the system used.

ABSTRAK

Kajian ini adalah untuk membangunkan satu kaedah baru mengenai Pengenalan Kerosakan menggunakan Penskalaan klasik. Proses Pemantauan adalah dari Kawalan Proses Statistik (SPC), alat statistik yang menggunakan kaedah statistik dan untuk mengawal proses, dengan mengulangi ukuran sampel atau untuk meramalkan keputusan. Ia juga membantu untuk menentukan sama ada proses ini berfungsi dengan betul atau tidak. Kawalan Proses Statistik (SPC) akan membentuk carta dengan data (carta kawalan) untuk menunjukkan hasilnya. Kajian ini akan membangunkan Proses Pemantauan Multivariat Statistik (MSPM) normal. Data akan berjalan seperti Proses Pemantauan Multivariat Statistik (MSPM) normal. Jadi dari teknik yang digunakan dalam kajian ini seperti Analisis Komponen Prinsipal, Kawalan Proses Statistik dan Proses Pemantauan Multivariat Statistik (MSPM), kajian ini akan membangunkan satu teknik yang paling cepat untuk mengesan kerosakan dalam sistem yang digunakan.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 BACKGROUND OF THE STUDY

Process Monitoring is an effective tool used to maintain control of a manufacturing operation, as well as demonstrating ongoing compliance to specifications.(Finepoint, 2003). It help to monitor some of critical points throughout the industrial process to increase productivity, maintain product quality, reduce downtime, prevent product loss or prevent accidents. It also will assist in controlling the product quality that relies on many factors such as temperature, pressure and flowrate. Chemical processes in industry require monitoring and management to maximize production and assist in providing safe product for consumption. With process monitoring, it's target to quickly identify deviation from the baseline performance. Other than that, there is Process Control which defines the active changing of the process based on the results of process monitoring. If the process monitoring tools have detected an out-of-control situation, the person responsible for the process makes a change to bring the process back into the control.

If problem happen, it can be solved using the problem resolution process. In this process, there are have 3 parts which are Fault Analysis, Fault Identification or Troubleshooting and Fault Resolution. For Fault Analysis, this process can hypothesize the cause of error and list the potential faults. For Fault Identification, it will eliminate and isolate the list of potential faults until offending fault is found, or we can mention that this fault will locate and identify the cause of the problem. And lastly is Fault Resolution is to apply the fix.

In Process Monitoring, there apply the Statistical Process Control (SPC) which is statistical tool for inspects random sample from a process. Statistical Process Control (SPC) also used to determine whether the process is in control or not. From the Statistical Process Control (SPC), it will forms charts with data (Control charts) to obtain the result of the process. For the control charts, there have two types of data which are variable data and attribute data. In Statistical Process Control (SPC), there have three types of fundamental which are understanding the process, understanding the causes of variation and elimination the sources of special cause variation. In understanding a process, usually the process using control charts, where to identify the variation that will be due into special causes. When the process capability is found lacking, it will be effort to determine causes of that variance. They usually used Ishikawa diagrams, designed experiments and Pareto charts as their tools.

1.2 PROBLEM STATEMENT

In Process Monitoring, there are many ways on detecting fault in the system process. For example, there have Statistical Process Control (SPC) and Principal Component Analysis (PCA). In the Statistical Process Control (SPC), we can get the results in control charts form. But there are limitations of Principal Component Analysis which it will take long step in detection the fault in the system. So that, this study will be develop the new Fault Identification method which is Classical Scaling.

1.3 RESEARCH PROBLEM

In this research, it will be more focusing on the variables of the process, using the concept of Process Monitoring and Statistical Process Control (SPC). Then, it will extend the process using Multivariate Statistical Process Monitoring (MSPM) versus Statistical Process Control (SPC), or using Classical Scaling (CMDS) which to compare the methods that will be detected the variables faster.

1.4 AIMS AND OBJECTIVES

1.4.1 Run the normal Multivariate Statistical Process Monitoring (MSPM)

1.4.2 Develop the new Multivariate Statistical Process Monitoring (MSPM)

1.4.3 Develop Fault Identification new method, which is Classical Scaling (CMDS)

1.5 SCOPE OF PROPOSED STUDY

In this research, it will propose a new technique on identify the fault in the process. It will be included many types of methods, to compare which one will be detect the variables faster. This study will be developed and run based on Matlab software.

1.6 EXPECTED OUTCOME

At the end of this research, the results will shows that using the Fault Identification new method, Classical Scaling (CMDS), that method will be detect the variables more faster than other methods.

1.7 SIGNIFICANT OF PROPOSED STUDY

When the new method, Classical Scaling (CMDS) will be develop, it help the system detect the faults and the variables more faster than the other methods. Therefore, it will produce the best quality of results in the system required.

1.8 CONCLUSION

This study is about to develop a new Fault Identification method which is Classical Scaling. This method is using the Multivariate Statistical Process Monitoring (MSPM) method. So, it will help to detect the fault faster rather use the Principal Component Analysis (PCA) method. It will help to make a better quality of product in the system. In the second chapter, the literature review will be classified into two sections that are fundamental of the Principal Component Analysis (PCA) and Multivariate Statistical Process Monitoring (MSPM). For the third chapter which is methodology, it will come out with the procedure of this study and the case study used in this study. For the forth chapter, which are results and discussions. It will come out with the results that will generate from the study and discussion about the responses to the results. Finally, the last chapter is conclusion which to conclude the study and to ensure that the study has achieved its objectives.

CHAPTER 2

LITERATURE REVIEW

2.1 Fundamental of Principal Component Analysis (PCA)

Principal Component Analysis (PCA) is one of the most common methods used by the data analysis to provide a condensed description and describe patterns of variation in multivariate data sets. Moreover, Principal Component Analysis is also able to retain meaningful information in the early axes whereas variation associated to experimental error, measurement in accuracy, and rounding is summarized in later axis (Gauch, 1982). Principal component analysis (PCA) is a well-established technique for monitoring and disturbance detection of multivariate process, as it enables variability assessment through dimensionality reduction. Principal component analysis is also a variable reduction procedure. When there have obtained the data on the number of variables and known that there is some disturbance in those variables, which means that some of the variables are correlated with one another. Because of the disturbance, it should be possible to reduce the observed variables into smaller number of principal component that will gives the most of the variance in the observed variables.

In the way to performing a principal component analysis, it is possible to calculate a score for each of subject on a given principal component. Then, the number of components extracted in a principal component analysis is equal to the number of observed variables being analyzed. However, in most analyses, only the first few components account for meaningful amounts of variance, so only these first few components are retained, interpreted, and used in subsequent analyses (such as in multiple regression analyses). This procedure is recommended to determine the minimum number of factors that will account for the maximum variance in the data in use in the particular multivariate analysis. This is a common technique for searching the patterns in the data that consist of high dimensions. From this technique, the data that will get is the standard deviation, covariance, eigenvectors and eigenvalues. The eigenvalues refer to the total variance explained by each factor. The standard deviation measures the variability of the data. The task of principal component analysis is to identify the patterns in the data and to direct the data by highlighting their similarities and differences.

In order to perform principal component analysis in an appropriate manner, we needs to subtract the mean from each of the dimensions of the data. The mean subtracted is the average across each dimension. This task is called data adjustment. The next task is to calculate the covariance matrix. The covariance can be computed only if the data is two dimensional. If the dimension of data is more than two, then the covariance is calculated or measured more than once. If the data is two dimensional, then the covariance matrix in principal component analysis is a square matrix with non diagonal elements in this matrix as positive elements. As the definition of eigenvalues, the calculation in principal component analysis involves the extraction of the total variance from each factor. The role of eigenvalues, which are then formed into vectors, is to provide the researcher with information about the patterns in the data. The principal component is referred to as that eigenvector. This has the highest eigenvalue. Once the eigenvectors are found from the covariance matrix, the next task is to sort the eigenvalues from highest to lowest. Thus, principal component analysis gives the original factors in terms of differences and similarities between the factors.

But sometimes, Principal component analysis is confused with factor analysis, because there are many important similarities between both of the procedures. Which mean that the both procedures are variables reduction methods that can be used to identify the observed variables that tend to get together empirically. Both procedures sometimes will be provide very similar results. But there are some important conceptual differences between principal component analysis and factor analysis that should be understood. This most related to the assumption of an underlying causal structure. Factor analysis assumes that the co-variation in the observed variables is due to the presence of one or more latent variables factors that exert causal influence on these observed variables. In contrast, principal component analysis wills no assumption about the underlying causal model. This is because principal component analysis is simply a variable reduction procedure that typically results in relatively small number of components that gives most of the variance in a set of observed variables.

2.2 Fundamental of Multidimensional Scaling

Multidimensional Scaling (MDS) is a method that represents measurements of similarity or dissimilarity among pairs of objects as distances between points of a low-dimensional multidimensional space (Borg & Groenen, 2005). Otherwise, Multidimensional Scaling (MDS) can be describes a family of techniques for the analysis of proximity data on a set of stimuli to reveal the hidden structure underlying the data (Steyvers, 2002). The proximity data can come from similarity judgments, identification confusion matrices, grouping data, same-different errors or any other measure of pairwise similarity. Multidimensional Scaling (MDS) is a set of data analysis techniques that display the structure of distance-like data as a geometrical picture.

Multidimensional Scaling can be classified according to whether the similarities data are qualitative is called as nonmetric MDS or quantitative known as metric MDS. The number of similarity matrices and the nature of the MDS model can also classify MDS types. This classification yields classical MDS which using one matrix and unweighted model, replicated MDS which used several matrices and unweighted model, and weighted MDS which using several matrices and weighted model. This method will give a graphical display of the structure of the data, one that is much easier to understand than an array of numbers and, moreover, one that displays the essential information in the data, smoothing out noise. This method also can be described by values along a set of dimensions that places these stimuli as points in a multidimensional space and that the similarity between stimuli is inversely related to the distances of the corresponding points in the multidimensional space. Groenen & Velden (2004) stated that Multidimensional Scaling (MDS) objective's to represent the dissimilarities as distances between points in allow dimensional space such that the distances correspond as closely as possible to the dissimilarities.

This method can be applied with different purposes. For example, by applying exploratory data analysis, which is by placing objects as points in a low dimensional space, the observed complexity in the original data matrix can often be reduced while preserving the essential information in the data. By a representation of the pattern of proximities in two or three dimensions, researchers can visually study the structure in the data. It also has been used to discover the mental representation of stimuli that explains how similarity judgments are generated. Sometimes, MDS reveals the psychological dimensions hidden in the data that can meaningfully describe the data. The multidimensional representations resulting from MDS are also often useful as the representational basis for various mathematical models of categorization, identification, and/or recognition memory (Nosofsky, 1992) or generalization (Shepard, 1987).

CHAPTER 3

METHODOLOGY

3.1 Procedure

3.1.1 Overall Framework in Multivariate Statistical Process Monitoring (MSPM) System

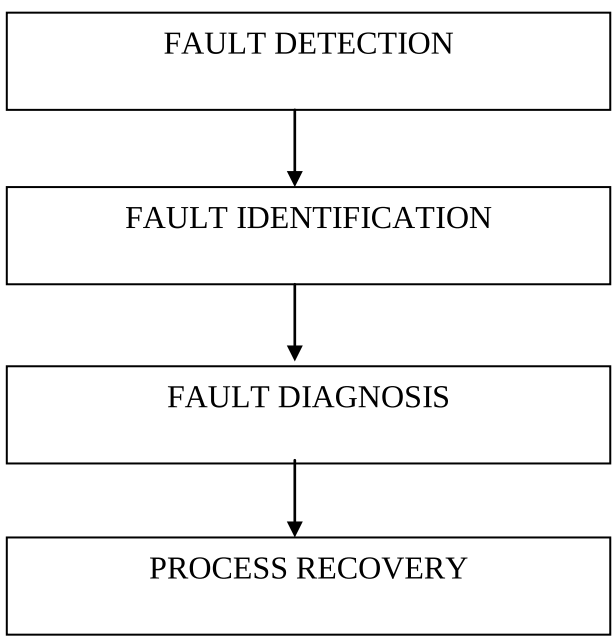

Figure 3.1.1: Overall Framework

In Multivariate Statistical Process Monitoring (MSPM) System, there are four main steps. First step is Fault Detection, which is to designate the departure of observed samples from an acceptable range using a set of parameters. For the second step is Fault Identification, which to identifying the observed process variables that are most relevant to the fault which is typically identified by using the contribution plot technique. For the third step which is Fault Diagnosis. This step is specially to determine the type of fault which has been significantly contributed to the signal. And lastly for forth step, Process Recovery to remove the causes that contribute to the detected fault.

3.1.2 Fault Detection and Fault Identification

This is the procedures of fault detection and fault identification to be complete.

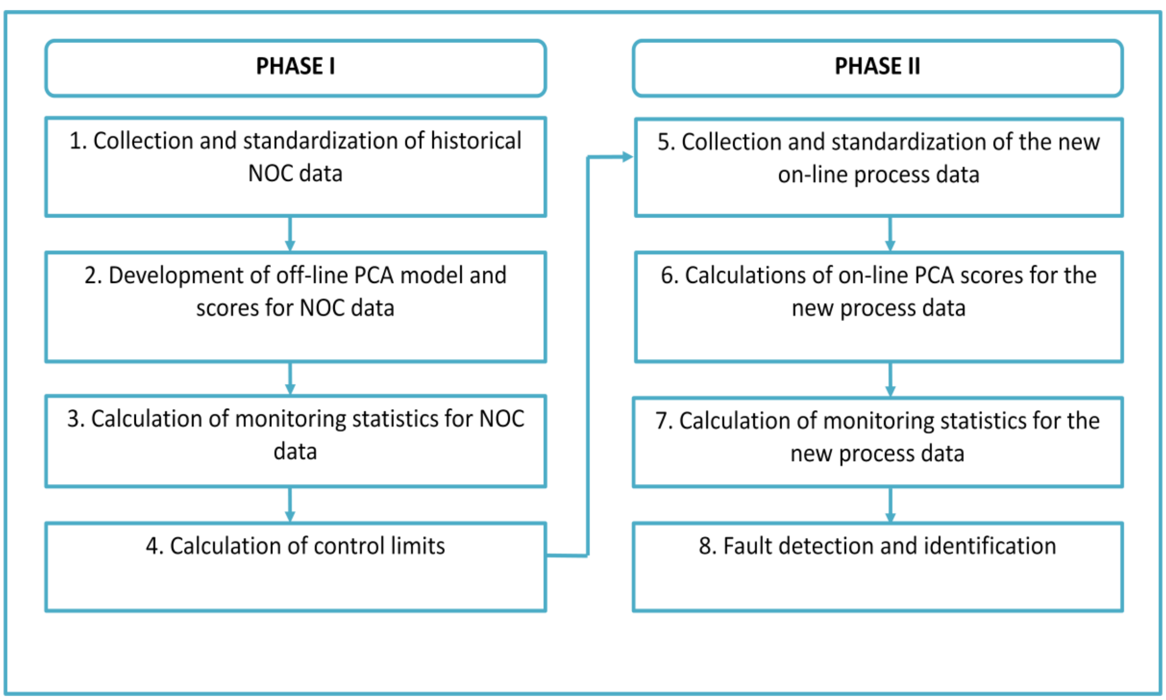

Figure 3.1.2 : Off-line Modeling and Monitoring (phase I) and On-line Monitoring (phase II)

Phase I, Off-line Modeling and Monitoring stage. In this phase, a set of normal operation condition data are identified off-line based on the historical process data archive. NOC simply implies that the process is operated at the desired setting condition and produces satisfactory products that meet the qualitative as well as quantitative specified standard (Martin et al., 1996). The data that we get will be standardized to zero mean and variance with respective to each of the variables because principle component analysis results depend on data scales.

In the second step, the development of principle component analysis model for the normal operation condition (NOC) data requires the establishment of a set of variance-covariance matrix, C_{mxm}.

$$\breve{x}_{j,i} = \frac{\left(x_{j,i} - \bar{x}_i\right)}{\sigma_i}$$

$$\mathrm{C} = \frac{1}{n-1}\breve{\mathrm{X}}'\breve{\mathrm{X}} = \begin{bmatrix} c_{1,1} & c_{1,2} & \cdots & c_{1,m} \\ c_{2,1} & c_{2,2} & \cdots & c_{2,m} \\ \vdots & \vdots & \vdots & \vdots \\ c_{m,1} & c_{m,2} & \cdots & c_{m,m} \end{bmatrix}$$

C is then transformed into a set of basic structures of eigen-based formula.

$$\mathrm{C} = \mathrm{V}\Lambda\mathrm{V}^{\mathrm{T}}$$

Finally, the principle component analysis model can be simply developed by:

$$\mathbf{P} = \breve{\mathbf{X}}\mathbf{V}$$

$$\begin{aligned} \mathbf{P} &= \begin{bmatrix} \mathbf{p_1} & \cdots & \mathbf{p_m} \end{bmatrix} \\ &= \begin{bmatrix} \breve{x}_{1,1}v_{1,1} + \cdots + \breve{x}_{1,m}v_{m,1} & \cdots & \breve{x}_{1,1}v_{1,m} + \cdots + \breve{x}_{1,m}v_{m,m} \\ \vdots & \cdots & \vdots \\ \breve{x}_{n,1}v_{1,1} + \cdots + \breve{x}_{n,m}v_{m,1} & \cdots & \breve{x}_{n,1}v_{1,m} + \cdots + \breve{x}_{n,m}v_{m,m} \end{bmatrix} \end{aligned}$$

The third step basically involves calculation of the Hotelling's T^2 and SPE monitoring statistics.

For T^2,

$$T_i^2 = \sum_{j=1}^{a} \frac{p_{i,j}^2}{\lambda_j}$$

For SPE

$$\begin{aligned}\tilde{\mathbf{E}} &= \breve{\mathbf{X}} - \hat{\mathbf{X}} \\ &= \breve{\mathbf{X}} - \mathbf{P}_a \mathbf{V}_{\mathrm{a}}^{\mathrm{T}} \\ &= \breve{\mathbf{X}} - \breve{\mathbf{X}} \mathbf{V}_a \mathbf{V}_a^{\mathbf{T}} \\ &= \breve{\mathbf{X}}\left(\mathbf{I} - \mathbf{V}_a \mathbf{V}_a^{\mathbf{T}}\right)\end{aligned}$$

$$SPE_i = \tilde{\mathbf{e}}_{\mathbf{i}} \tilde{\mathbf{e}}_{\mathbf{i}}^T$$

For the final step in phase I, it will deals with developing the control limits for both of the statistics.

$$T_\alpha = \frac{A(n-1)}{(n-A)} F_{A,n-A,\alpha}$$

$$SPE_\alpha = \theta_1 \left(\frac{z_\alpha \sqrt{2\theta_2 h_0^2}}{\theta_1} + \frac{\theta_1 h_0 (h_0 - 1)}{\theta_1^2} + 1\right)^{\frac{1}{h_0}}$$

In phase II, is for On-line Monitoring. In this phase, the step 5 until step 7 is similar to the step 1 until 3 in the phase I. Until step 8 which are the last step, there are two main operations that have to be conducted separately, which are fault detection and fault identification. For fault detection, it is a fault situation that regards as a result of an occurrence of a special event that is not in conformance to the common cause nature. A fault situation will be declared if either of the monitoring statistics exceeding its respective control limit for a pre-defined successive number of samples. In this phase also, the data will be more focusing on the abrupt fault and the incipient fault. After that, fault identification will be based on the contribution plot which the variables have given the most contribution to the fault.

3.2 Case Study

The multivariate statistical process control approaches are based on noncausal models built from historical process data using multivariate latent variable methods such as PCA and PLS. The faults are then detected by referencing future data against these covariance models, and isolation is attempted through examining contributions to the breakdown of the covariance structure. There are major differences between these approaches arising mainly from the different types of models and data utilized to build them. Each of them has clear, but complementary, strengths and weaknesses. These are discussed using simulated data from a CSTR process shown below.

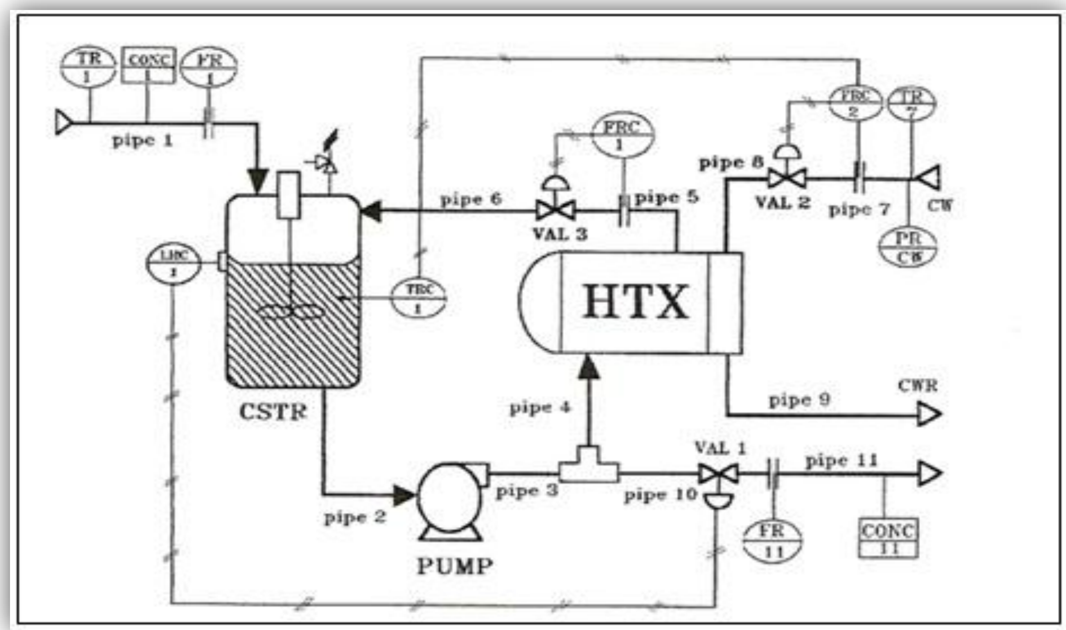

Figure 3.2a: CSTR system (Zhang, 2006)

Table 3.2b : List of Abnormal Operation in the CSTR System

Fault Cases	**Fault Causes**
1	Pipe 1 blockage
2	External feed-reactant flow rate too high
3	Pipe 2 or 3 is blocked or pump fails
4	Pipe 10 or 11 is blocked or control valve 1 fails low
5	External feed-reactant temperature abnormal
6	Control valve 2 fails high
7	Pipe 7,8 or 9 is blocked or control valve 3 fails low
8	Control valve 1 fails high
9	Pipe 4,5 or 6 is blocked or control valve 3 fails low
10	Control valve 3 fails high
11	External feed-reactant concentration too low

Table 3.2c : List Of Variables In The CSTR System For Monitoring

Process		
No	**Variables**	**Variables Names**
1	V1	Tank temperature
2	V2	Tank level
3	V3	Feed temperature
4	V4	Inlet flow rate
5	V5	Recycle flow rate
6	V6	Outlet flow rate
7	V7	Cooling water flow rate
8	V8	Product concentration
9	V9	Feed concentration
10	V10	Heat exchanger entrance pressure
Instruments		
11	V11	Controller 1
12	V12	Controller 2
13	V13	Controller 3

In Figure 3.2a, it showed the case study that used in this study. In this case study, the researchers have listed the abnormal operation that might be happen in the CSTR system in Table 3.2b. There are many fault that might be happen in this case study such as the pipe are blockage, external feed-reactant too high or the control valve are fails. In Table 3.2c, there are list of the variables occurred in the CSTR system to monitor.

CHAPTER 4

RESULT AND DISCUSSION

4.1 Normal Operation Condition (NOC) Data

When start run the phase I, the first data that will be get is the normal operation condition (NOC) data. this data will be as the reference data when to detecting the variable that contribute to the fault. The data that will be get as below which are T^2 Statistics versus Observations data and SPE Statistics versus Observations data.

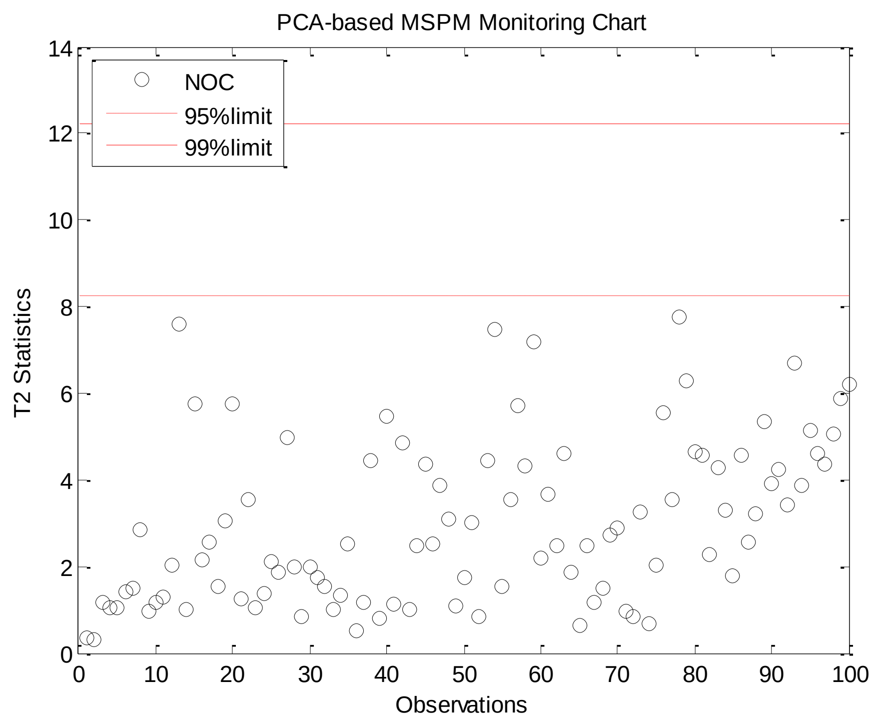

Figure 4.1a: T^2 Statistics versus Observations

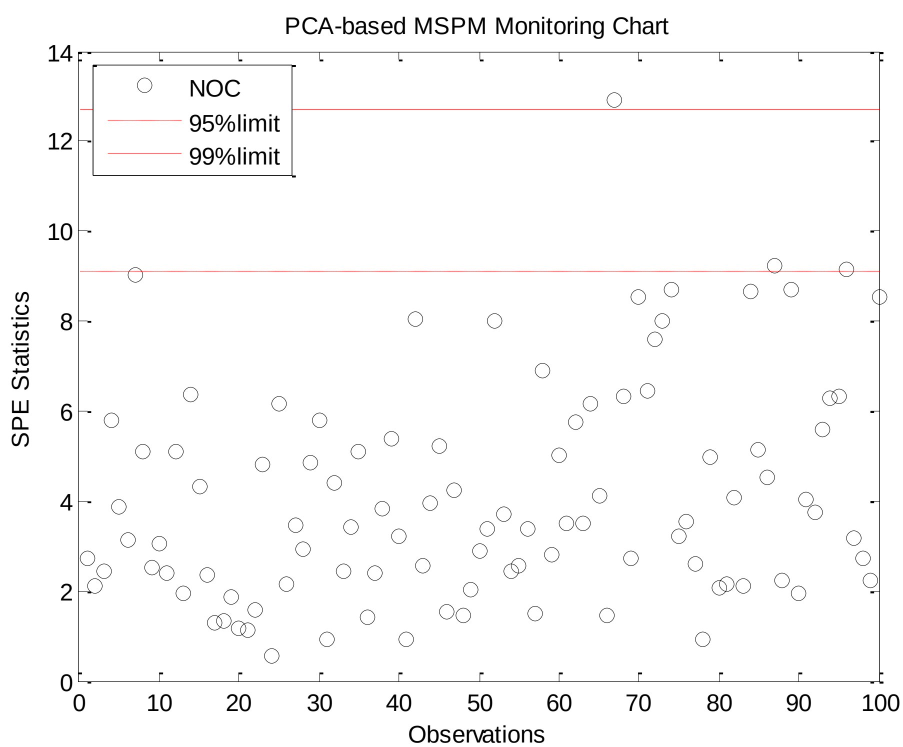

Figure 4.1b: SPE Statistics versus Observations